Computational Statistics and Data Analysis 57 (2013) 570–588

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Stochastic models for multiple pathways of temporal natural history on co-morbidity of chronic disease

Amy Ming-Fang Yen [a,b], Hsiu-Hsi Chen [b,*]

[a] School of Oral Hygiene, College of Oral Medicine, Taipei Medical University, Taipei, Taiwan

[b] Graduate Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 24 July 2010
Received in revised form 2 December 2011
Accepted 7 July 2012
Available online 27 July 2012

Keywords:
Chronic disease
Co-morbidity
Latent variable
Markov process

ABSTRACT

Chronic diseases frequently co-occur in individuals. Susceptibility to co-morbidity, the temporal sequence and the transition rates governing the development of co-morbid diseases are often hidden or partially observable. To tackle these thorny issues we developed a series of co-morbidity stochastic models with latent variables to estimate the true proportions of susceptibility, temporal sequence, and transition rates. We begin with a bivariate co-morbidity model for two chronic diseases, then extend to a trivariate co-morbidity model for three chronic diseases, and to a generalized high-order co-morbidity model to accommodate more than three chronic diseases. To illustrate our approach we fitted the proposed model with data from a population-based health check-up for hypertension, diabetes mellitus (DM), and overweight in Matsu.

Compared with 3.93% of co-morbidity directly estimated from empirical data, approximately 12% (10%–14%) of participants have the potential of developing both hypertension and DM from the underlying population. Hypertension prior to DM was 74% (54.10%–93.77%) of these subjects susceptible to co-morbidity. Those who developed DM first had a higher likelihood of having hypertension (65.85 per 100 person-years; 95% CI: 15.61–116.09) compared with those with hypertension first and DM later (36.37 cases per 100 person-years; 95% CI: 14.57–58.18). Gender, smoking, and alcohol drinking modeled by incorporating them as covariates with proportional hazards form had impacts on different parameters of interest. The deviance statistics, indicating a lack of statistical significance (p values were 0.26 for the bivariate model) for the model without covariates and for the model with covariates (all p values > 0.05), suggest a satisfactory model fit. However, the trivariate co-morbidity model had poorer fit than the bivariate co-morbidity model.

Our proposed co-morbidity stochastic latent variable models can tackle the problem of underestimating the proportion of susceptibility to co-morbidity, giving a clue to the temporal sequence of a constellation of co-morbid diseases, and quantifying the incidence rates of each disease and the corresponding transitions rates between co-morbid diseases. The generalized high-order co-morbidity model can be extended to model the complex pathway of high dimension of chronic diseases in the clinical field provided the dataset is sufficiently large.

* Correspondence to: Graduate Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Room 533, No. 17, Hsuchow Road, Taipei, 100, Taiwan. Tel.: +886 2 33668033; fax: +886 2 23587707.
E-mail address: chenlin@ntu.edu.tw (H.-H. Chen).

0167-9473/$ – see front matter © 2012 Published by Elsevier B.V.
doi:10.1016/j.csda.2012.07.009

1. Introduction

Chronic diseases, such as type 2 diabetes, hypertension, hyperlipidemia, hyperuricemia and overweight, frequently co-occur in individuals. A population-based survey found the proportion of co-existing chronic disease in adults with diabetes mellitus (DM) was 59% and 42% for hypertension and hyperlipidemia, respectively (Chen et al., 2004). From a public health viewpoint, instead of focusing on patients, it is important to elucidate co-morbid disease rates derived from the underlying population. This is highly justified by the recently proposed metabolic syndrome including multi-dimensional indicators (hypertension, overweight, hyperglycemia, low level of high-density lipoprotein and hypertriglyceride), because the elucidation of the magnitude and temporal sequence of these co-existing multi-dimensional components provides a good guidance for designing population-based prevention programs (Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults, 2001).

Attempts have been made at evaluating co-morbidity due to chronic diseases, but four key issues have not been fully addressed yet. Firstly, in many studies the estimated frequency of co-morbidity is based on the prevalence rate at a specific time with a cross-sectional survey (Al-Mahroos et al., 2000; Briganti et al., 2003; Kannel et al., 1991; Mitchell et al., 1999; Onal et al., 2004; Tai et al., 1991; Wannamethee et al., 1998). Information based on a cross-sectional survey makes it difficult to distinguish those who are at risk of developing co-morbidity, but have not been observed at the time of the survey, so called right-censoring, from those who are not at risk (non-susceptibility). The lack of knowledge of non-susceptibility in the population tends to underestimate the co-morbidity rate.

Secondly, the elucidation of the temporal sequence of multiple chronic diseases, a potentially thorny issue, is hampered by partially observed data because several co-morbid diseases have been detected at a one-shot survey or even repeated cross-sectional surveys. Elucidating the temporal sequence of alternating chronic diseases is very informative in designing a population-based intervention program for reducing the progression of chronic disease such as metabolic syndrome.

Thirdly, as the exact onset of each chronic disease or subsequent progressions to other co-morbid diseases under consideration is often impossible to pinpoint, especially when more co-morbid diseases have been involved, it is also difficult to estimate transition rates related to the time to the onset of each disease and also the time intervals between the occurrence of two diseases by simply relying on the traditional survival analysis.

Fourthly, relevant covariates such as life-style factors affecting susceptibility to co-morbidity and transitions between multiple events under consideration further complicate the development of such sophisticated models.

These points strongly suggest a complex statistical model is required. We show the relevant data dealing with these thorny issues and illustrate the pathway for the bivariate co-morbidity model in the next section.

2. Illustration of data on the pathway of bivariate co-morbidity model

2.1. Data

As health check-ups for chronic disease are popular, health policy-makers are interested in assessing the frequency of co-morbidity and the temporal sequence of occurrence of chronic diseases after a health check-up in order to implement health planning for the prevention of chronic diseases. We used data from a population-based screening program in Matsu (a small island located in the Taiwan Strait) in which 3571 registered residents, aged 30 years or above, were invited to screening for chronic diseases between 1995 and 1999. The attendance rate was 61.6% (2201 individuals) with information on demographic and life-style factors, diet, family cancer history and personal medical details being obtained via a structured questionnaire. Serological measurements were also taken during the program. Of these serological markers, we are interested in four continuous variables, systolic blood pressure, diastolic pressure, fasting blood glucose, and body mass index (BMI). It should be noted that although these four variables can be modeled as an interval scale property, disease and co-morbid disease rather than the dynamic change of four continuous variables are of great interest in our study. Hypertension, self-reported in some cases, was defined for new diagnoses as systolic blood pressure $\geq$140 mmHg or diastolic blood pressure $\geq$90 mmHg. Similarly, type 2 diabetes was either self-reported or defined as fasting blood glucose $\geq$126 mg/dL. Overweight was defined as body mass index $\geq$25 kg/m^2.

2.2. Pathway for bi-variate co-morbidity model

We illustrate a bivariate co-morbidity model by defining all possible clinical pathways in Fig. 1 for the example of hypertension (HTN) and diabetes mellitus (DM) to account for the motivation of doing this study. The upper panel is the pathway for those who are not at risk of developing co-morbidity and only have, at most, a single disease entity, either hypertension (pathway 1a) or DM (pathway 1b). The lower two panels are the pathways for those who are susceptible to co-morbidity with either hypertension preceding DM (pathway 2) or DM prior to hypertension (pathway 3).

Table 1 shows empirical data relating to our bi-variate co-morbidity model (with hypertension and type 2 diabetes) for individual data ($l = 1, \ldots, 1489$) on four rounds ($x = 1, 2, 3, 4$) of screening in Matsu between 1995 and 1999 and the corresponding time, $t_{l,x}$, with the disease status denoted by $y_{l,x}$. Table 2 shows data by transition histories together with the corresponding observed counts, and possible pathways. The main motivations for developing a series of stochastic